

Extraction de motif de mouvement de la foule par le flot optique

H. Chebi, et D. Acheli

Faculté des hydrocarbures et de la chimie, Université M'hamed Bougara Boumerdès
Avenue de l'Indépendance 35000 Boumerdès, Algeria

chebi.hocine@yahoo.fr; dacheli2000@yahoo.fr.

Résumé — Dans ce travail, nous nous intéressons à l'analyse d'extraction de motifs de mouvement par la technique de flot optique, dans le but d'analyser le comportement de la foule. Ces scènes se caractérisent par la présence d'un grand nombre de personnes dans le champ de vision des caméras. Le problème majeur est l'élaboration et d'utilisation d'une technique sans modélisation de l'arrière-plan pour détecter les mouvements de la foule. Par la suite une étape de détection d'anomalies par la technique des réseaux de neurones artificiels (RNA). Nous présentons dans cet article une comparaison de l'approche proposée pour la détection des mouvements dans des scènes très denses et celle d'utilisation de modèle social de force. Pour plus de robustesse et d'efficacité, nous avons introduit la routine permettant d'élimination des ombres.

Mots Clés — Analyse visuelle, analyse de foules, élimination d'ombres, vidéosurveillance intelligentes, classification supervisée, réseaux de neurones (RNA), classification, flot optique.

I. INTRODUCTION

L'essor des systèmes d'acquisition et de traitement de la vidéo joue un rôle important dans la vie quotidienne. Cette importance évolue proportionnellement avec nos besoins d'automatiser le processus d'extraction de l'information depuis la vidéo.

L'étude de foules humaines est une discipline complexe. Elle regroupe à la fois la notion de comportement individuel et collectif, chacun en soi étant relié à des mécanismes divers selon les individus et les sociétés ; d'où la nécessité de déceler des comportements anormaux afin de prévenir des actes dangereux portant atteinte à la société [1], [2]. Cette étape est nécessaire afin de permettre une analyse précise des événements qui se déroulent dans une scène (interaction entre deux personnes, détection de comportements dangereux, vitesse excessive, chute,...etc.). De plus, avec l'augmentation du nombre de caméras ainsi que de leur résolution, il est nécessaire de développer des solutions très rapides afin de pouvoir respecter des contraintes temps-réel avec une

intervention minimum des opérateurs (gestion du plus grand nombre possible d'anomalies).

L'approche proposée dans cet article pour la détection d'anomalies se distingue des approches existantes [1], [2], [3] par son mécanisme dynamique permettant de détecter d'une manière automatique les anomalies (cas d'un comportement normal et anormal). Cette dernière se décompose en trois niveaux : niveau bas (estimation du flux optique), niveau intermédiaire (construction du modèle de magnitude, et utilisation des techniques de traitement d'image) et niveau sémantique (notification des opérateurs).

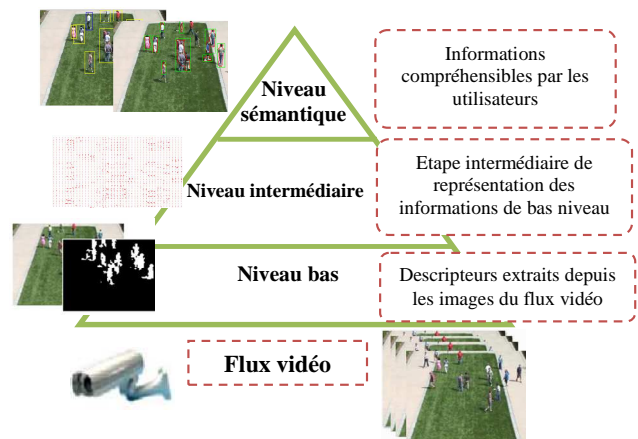


Fig. 1. Illustration de l'approche globale proposée pour l'analyse du comportement humain dans une scène dense.

Le but de l'approche utilisée est la détection d'anomalies dans des scènes très denses. Les différentes anomalies sont détectées d'une manière automatique avec apprentissage par l'approche du réseau de neurone. Pour plus de robustesse et d'efficacité, nous avons introduit et proposés la routine permettant l'élimination des ombres [4], [5].

Le reste du papier est organisé comme suit: dans la section II nous présentons un bref état de l'art des approches ainsi que les travaux relatifs à ce domaine. Comme nous illustrons les

techniques utilisées afin de palier à certains problèmes rencontrés en littérature. Dans la section III nous présentons la formulation mathématique des différentes méthodes utilisées afin de détecter des anomalies de type course et marche, dans une scène dense. Des résultats sont discutés dans la section IV. Nous terminons par une conclusion sur le travail réalisé et des perspectives pour des travaux futurs.

II. ETAT DE L'ART

Alors que la dérivée temporelle quantifie la variation de l'aspect de chaque pixel considéré individuellement dans le cas de la détection de mouvements, le flot optique est modélisée par un champ de vecteurs à deux dimensions représentant la projection sur le plan image du mouvement réel observé (3D). Dans cette optique, de nombreuses méthodes [6], [7], [8] ont été proposées depuis l'article précurseur de Horn et Schunck [9] dans le but d'améliorer l'implémentation de cette dernière. Dans [10], neuf algorithmes sont étudiés et comparés selon des critères de précision et de densité du champ de vecteurs obtenu, mais aucune mention n'est faite sur la complexité algorithmique. Les travaux de [11] permettent de combler cette lacune en mesurant les rapports (précision)/(temps de calculs) de ces méthodes.

La segmentation est une approche qui consiste à découper l'image de façon successive en estimant le mouvement apparent dominant des zones non étiquetées puis en détectant les zones non conformes à ce modèle de mouvement [12]. Le principe de cette méthode repose sur la prise en compte d'une image globale sur laquelle le mouvement dominant et les zones conformes à ce mouvement formant une première région de la partition sont estimées. Ce procédé d'estimation et de détection est itéré sur les zones non conformes jusqu'à ce que l'ensemble des pixels soient classés [13], [14].

La méthode d'apprentissage en vue d'une phase de classification génère une fonction qui fait correspondre à une image reçue en entrée une étiquette spécifique. Il existe plusieurs méthodes d'apprentissage en littérature, telles que les arbres de décision [15], les réseaux de neurones [16], la logique floues, l'AdaBoost (Adaptive Boosting) [17], ou les machines à vecteurs de support (SVM) [18]...etc.

Les méthodes de poursuite proposent de reconnaître et localiser au cours du temps des objets présents dans une séquence temporelle d'images [19], [20]. Dans le cadre des foules humaines, elles trouvent un intérêt particulier dans la vidéo surveillance où le suivi des individus permet de contrôler automatiquement les allées et venues dans un espace. Tout comme la reconnaissance à partir d'une image, la poursuite peut s'appuyer sur des propriétés graphiques telles

que les couleurs ou les contours [21], [22]. La dimension temporelle ajoutée permet en outre de supposer une continuité de la présence et de la position des objets dans la scène, malgré les occultations. La consistance temporelle et spatiale des caractéristiques suivies peut dans certains cas s'obtenir à l'aide de méthodes de partitionnement (clustering) [23].

Dans notre travail, nous proposons l'utilisation de la technique de détection de mouvement par le flot optique (technique de Horn et Schunck). Cette dernière, permet de détecter des groupes qui se déplacent dans la même direction et d'extraire les motifs de mouvement. L'avantage majeur est que c'est une méthode non modélisable [1], consistant à détecter le mouvement par le calcul en tout point de l'image d'une quantité mathématique qui est fonction de l'intensité ou de la couleur de l'ensemble des pixels et qui est censée refléter l'importance du mouvement visible dans la scène. Nous proposons par la suite la segmentation par regroupement des régions dans le but de fournir un découpage plus précis des frontières des régions, qui serviront à la phase de classification supervisée par réseau de neurone. Nous appliquerons une méthode permettant d'annihiler les effets d'ombre, afin d'avoir une bonne détection des personnes. Nous proposons par la suite l'amélioration des résultats obtenus en ajoutant la technique d'élimination automatique d'ombres.

III. FORMULATION MATHÉMATIQUE

A. Extraction des mouvements

Cette tâche est accomplie en calculant le flux optique par l'algorithme de Gunnar Farneback [24] à partir de deux images successives.

L'application du flux optique nous permet d'obtenir l'ensemble des vecteurs de mouvement suivant:

$$V_{i,t} = \{V_1, \dots, V_N | V_i = (x_{i,t}, y_{i,t}, m_{i,t}, \theta_{i,t})\} \quad (1)$$

Où :

- $x_{i,t}$ et $y_{i,t}$ sont les coordonnées de l'origine du vecteur $V_{i,t}$.
- $m_{i,t}$ est la magnitude du vecteur de mouvement $V_{i,t}$.
- $\theta_{i,t}$ est l'orientation du vecteur de mouvement $V_{i,t}$.

B. Classification d'un groupe de personnes par RNA

Le réseau de neurones est utilisé pour classifier le comportement de foules. La classification est faite pour chaque séquence d'images afin de déceler les différents comportements d'un groupe de personnes. Le classificateur est basé sur un ensemble de réseau de neurones contenant trois couches. Chaque type de comportement est détecté par un réseau individuellement qualifié. La structure de chaque

réseau est définie de telle sorte que le nombre de neurones dans la couche d'entrée est égal au nombre de paramètres d'entrée, et le nombre de neurones dans la couche cachée est déterminé expérimentalement. Dans ce travail où deux classes sont considérées, à savoir le comportement normal et anormal, chaque type de comportement est représenté par un neurone dans la couche de sortie. Les neurones dans la couche cachée sont représentés par une fonction sigmoïde. Le modèle est représenté l'équation suivante :

$$e_k^{(i)} = \sum_j w_{j,k}^{(i-1)} s_j^{(i-1)} + b_k^{(i-1)} \quad (2)$$

En sortie, nous avons $s_k^{(i)} = a_k^{(i)}(e_k^{(i)}) = \frac{1}{1+\exp(-e_k^{(i)})}$ qui est la fonction d'activation appliquée dans le réseau. La valeur $b_k^{(i-1)}$ est un biais ajouté à l'entrée du k-ième neurone. La fonction d'erreur, qui représente la distance euclidienne entre la sortie du réseau et la cible est donnée comme suit :

$$E = \frac{1}{2} \|y - c\|_2^2 = \frac{1}{2} \sum_{i=1}^m (y_i - c_i)^2 \quad (3)$$

Il faut maintenant minimiser la moyenne des erreurs données par la fonction E sur l'ensemble des données fournies en entrées :

$$E_{moyenne} = \frac{1}{N} \sum_{t=1}^N E_t \quad (4)$$

Avec N est le nombre de couples donné et E_t représente la t-ième erreur d'apprentissage.

C. Elimination d'ombre

L'étape de suppression des ombres a été accomplie par un passage de l'espace colorimétrique RGB vers l'espace HSV (pour *Hue / Saturation / Value*), tout en éliminant la composante luminance V. Cet espace met en évidence les propriétés chromatiques grâce à ses composantes H, S et V.

La chromaticité du pixel (x, y) peut être obtenue par la normalisation de la couleur, comme par exemple pour le canal rouge :

$$C(r, x, y) = \frac{I(r, x, y)}{I(r, x, y) + I(g, x, y) + I(b, x, y)} \quad (5)$$

Où $I(r, x, y)$, $I(g, x, y)$, $I(b, x, y)$ sont respectivement les intensités des canaux rouge, vert et bleu.

Il est de même pour les deux autres canaux vert et bleu. La chromaticité d'un pixel est alors :

$$C(x, y) = (C(r, x, y), C(g, x, y), C(b, x, y)) \quad (6)$$

La différence de chromaticité ΔC , entre la chromaticité courante $c'(x, y)$ du pixel et sa chromaticité $c(x, y)$ dans le modèle de l'arrière-plan permet de mesurer la distance de Mahalanobis $D(x, y)$ définie formellement comme suit :

$$D(x, y) = \sqrt{\Delta C(x, y)^T S^{-1} \Delta C(x, y)} \quad (7)$$

Où $S(x, y)$ est la matrice de covariance de la chromaticité du pixel (x, y) .

Alors, pour chaque nouvelle image et pour chaque pixel, la distance de Mahalanobis est calculée.

Si cette distance est supérieure à une constante (qui varie de 3 à 5) le pixel est attribué au premier plan, dans le cas contraire il est considéré comme ombre, et donc attribué au fond.

D. Détection d'évènements

Dans cette section, nous décrivons la détection d'anomalies dans une scène dense. Les scénarii choisis font partie des évènements décrits dans la vidéo [25]. Les comportements décelés est course et marche :

L'idée principale consiste à calculer la magnitude moyenne des vecteurs de mouvement dans chaque image. Une magnitude élevée signifie l'évènement course tandis qu'une magnitude faible signifie l'évènement marche.

$$M_{moy} = \frac{\sum_{i=1}^{n \times m} mag(i)}{n \times m} \quad (8)$$

Avec n et m le nombre des points de mouvement dans l'image, et $mag(i) = \sqrt{Rel_i^2 + Img_i^2}$ est la magnitude de chaque points de mouvement.

Où Rel et Img sont les composantes réelles du et imaginaires du vecteur de mouvement.

IV. RESULTATS

Le principe de l'approche utilisée est l'obtention de la magnitude du vecteur de mouvements qui exprime un vecteur de flux optique dans un repère cartésien (fig.2). Le point $P(X_i; Y_i)$ est la position du point d'intérêt i dans l'image t , tandis que $Q(X_i; Y_i)$ est sa position dans l'image $t+1$. Nous pouvons facilement calculer la distance entre ces deux points grâce à la distance euclidienne :

$$M_i = \sqrt{(Q_{xi} - P_{xi})^2 + (Q_{yi} - P_{yi})^2} \quad (9)$$

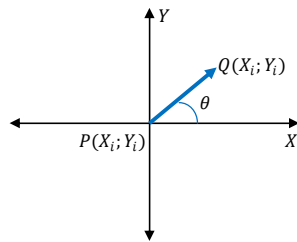


Fig. 2. Représentation d'un vecteur de flux optique dans un repère cartésien.

Une scène de foule contient un nombre variable de personnes. Les personnes sont souvent occultées et il est difficile de distinguer leurs formes. Il n'est donc pas intéressant d'utiliser une méthode d'extraction de l'arrière-plan dont l'utilité principale est la détection des personnes. Le flux optique s'avère très utile dans ce genre de scènes car il permet de détecter des groupes qui se déplacent dans la même direction et d'extraire les motifs de mouvement (voir fig.3).

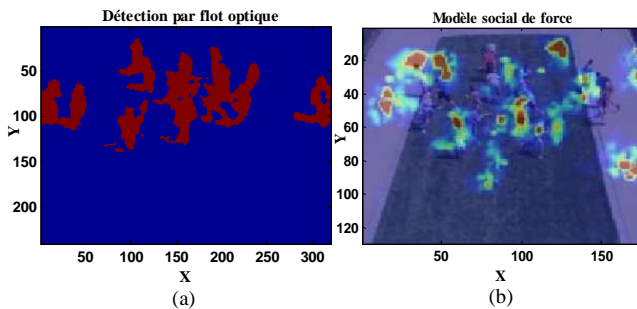


Fig. 3. Exemple de détection de motif de mouvement de la scène de foules, (a) Approche proposée, (b) Approche de Ramin Mehan [26].

Par la suite nous calculons la magnitude moyenne de chaque séquence d'images, dans le but de déterminer l'événement course et marche (fig.4). Ces événements peuvent être identifiés en utilisant les vecteurs du flux optique. La détection se fait en utilisant les réseaux de neurones avec comme caractéristiques la vitesse moyenne du mouvement. Les résultats de l'approche proposée sont représentés par la fig.4.

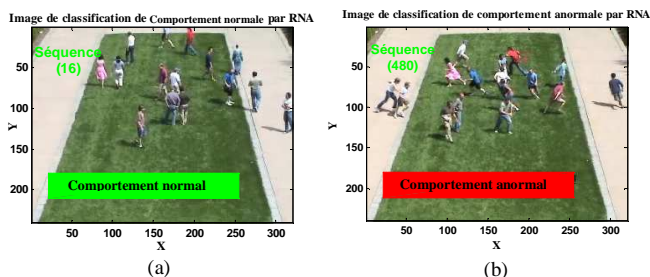
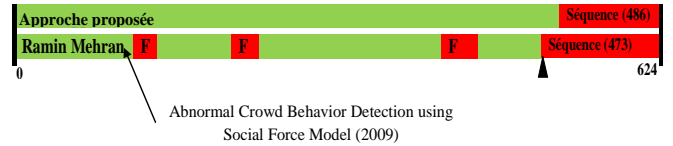


Fig. 4. Exemple d'analyse de comportements par réseau de neurone, (a) comportement normal, (b) comportement anormal.

Dans le travail élaboré, nous supposons que le nombre de personnes dans un groupe occulté n'est pas limité. D'ailleurs, nous comparons nos résultats à la méthode (fig.5), à savoir, le

modèle social de force [26], On remarque que l'approche possède des fausses détections plutôt que de manquer de réelles détections « présence de fausses comportement anormal », cette méthode présente la vision d'ordinateur basée sur l'advection de particules pour détecter et localiser le comportement anormal de foule en utilisant le modèle social de force.



❖ F : Représente les fausses alertes détectées.

Fig. 5. Moments de détection d'anomalies (vert : Evénements normaux, rouge : Evénements anormaux).

Nous présentons quelques résultats illustratifs dans les (fig. 6) exprimant le cas d'élimination d'ombre. Nous notons que notre approche décrite certaines avantage comme il présente une contribution positive pour la détection des motifs de mouvement dans un environnement complexe. Aussi l'inconvénient d'être très gourmande en puissance de calcul car elle nécessite l'estimation d'une série temporelle pour chaque bloc d'images et ce, à chaque instant de la vidéo. Car aussi le flot optique il n'occupe pas une grande surface de détection 84 % de mouvement, par apport à l'advection de particules de modèle social force qui serre à localiser une grande surface plus de 96 %.

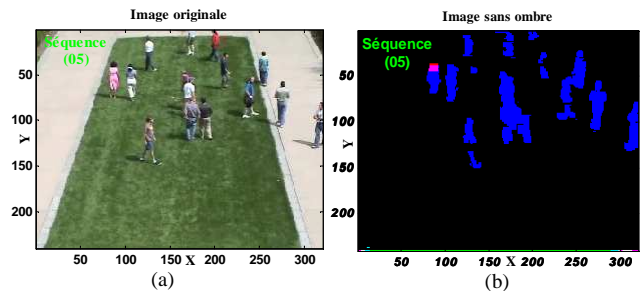


Fig. 6. Résultats de détection avant et après la suppression de l'ombres. (a) Images originale, (b) Elimination d'ombres.

V. CONCLUSION

À l'issu de ce travail, nous avons présenté notre approches pour détecter des comportements anormaux dans des scènes de foule. En utilisant le flot optique. Les résultats de notre méthode, indique que la méthode est efficace dans la détection et la localisation des comportements anormaux dans la foule, mais l'approche utilisée quant à elle, atteint un débit de 3 images par seconde sur un processeur d'ordinateur portable de faible puissance de marque Intel pentium 2.16 GHz, simulé sous MATLAB « R2014a ». Comme perspectives, nous proposons d'améliorer les performances de la méthode utilisée en introduisant et implémenté ces techniques sur un carte FPGA.

REFERENCES

- [1] Shu Wang; Zhenjiang Miao, "Anomaly detection in crowd scene," *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, vol., no., pp. 1220, 1223, 24-28 Oct. 2010.
- [2] Pathan, S.S.; Al-Hamadi, A.; Michaelis, B., "Crowd behavior detection by statistical modeling of motion patterns," *Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of*, vol., no., pp.81, 86, 7-10 Dec. 2010.
- [3] Ko, T., "A survey on behavior analysis in video surveillance for homeland security applications," *Applied Imagery Pattern Recognition Workshop, 2008. AIPR '08. 37th IEEE*, vol., no., pp.1,8, 15-17 Oct. 2008.
- [4] Jing Li; Geng Wang, "A shadow detection method based on improved Gaussian Mixture Model," *Electronics Information and Emergency Communication (ICEIEC), 2013 IEEE 4th International Conference on*, vol., no., pp.62, 65, 15-17 Nov. 2013.
- [5] Huihuan Qian; Xinyu Wu; Yongsheng Ou; Yangsheng Xu, "Hybrid algorithm for segmentation and tracking in surveillance," *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*, vol., no., pp.395,400, 22-25 Feb. 2009.
- [6] Shi, J.; Tomasi, C., "Good features to track," *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, vol., no., pp.593,600, 21-23 Jun 1994.
- [7] Atcheson, W. Heidrich, I. Ihrke, "An evaluation of optical flow algorithms for background oriented schlieren imaging." *Experiments in Fluids*, vol. 46, pp. 467-476, 2009.
- [8] Burt, P.J.; Adelson, E.H., "The Laplacian Pyramid as a Compact Image Code," *Communications, IEEE Transactions on*, vol.31, no.4, pp.532,540, Apr 1983.
- [9] Horn, Berthold K., and Brian G. Schunck. "Determining optical flow." *1981 Technical Symposium East*. International Society for Optics and Photonics, 1981.
- [10] Barron, John L., David J. Fleet, and Steven S. Beauchemin. "Performance of optical flow techniques." *International journal of computer vision*, 43-77, 12:1:1994.
- [11] Liu, Hongche, al. "Accuracy vs efficiency trade-offs in optical flow algorithms." *Computer vision and image understanding* 271-286, 72.3:1998.
- [12] Demonceaux, Cédric. *Etude du mouvement dans les séquences d'images par analyse d'ondelettes et modélisation markovienne hiérarchique. Application à la détection d'obstacles dans un milieu routier*. Diss. Université de Picardie Jules Verne, 2004.
- [13] M. Irani, B. Rousso, S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In Proc. European Conference on Computer Vision, pages 282–287, Santa Margherita Ligure, Italy, Mai 1992.
- [14] S. Ayer, P. Schroeter, J. Bigun. Segmentation of moving objects by robust motion parameter estimation over multiple frames. In Proc. European Conference on Computer Vision, pages 316–327, Stockholm, Suède, Mai 1994.
- [15] Lynne GREWE, Avinash C. KAK : Interactive learning of a multiple-attribute hash table classifier for fast object recognition. *Computer Vision and Image Understanding (CVIU)*, 61(3):387–416, 1995.
- [16] Henry A. ROWLEY, Shumeet BALUJA, Takeo KANADE: Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 20:23–38, 1996.
- [17] Paul VIOLA, Michael JONES, Daniel SNOW: Detecting pedestrians using patterns of motion and appearance. In International Conference on Computer Vision (ICCV), pages 734–741, 2003.
- [18] C.P. PAPAGEORGIOU, M. OREN, T. POGGIO : A general framework for object detection. In International Conference on Computer Vision (ICCV), pages 555–562, 1998.
- [19] Pierre F. GABRIEL, Jacques G. VERLY, Justus H. PIATER et André GENON: The state of the art in multiple object tracking under occlusion in video sequences. In In Advanced Concepts for Intelligent Vision Systems (ACIVS), 2003, pages 166–173, 2003.
- [20] Serajeh, R.; Faez, K.; Ghahnavieh, A.E., "Robust multiple human tracking using particle swarm optimization and the Kalman filter on full occlusion conditions," *Pattern Recognition and Image Analysis (PRIA), 2013 First Iranian Conference on*, vol., no., pp.1,4, 6-8 March 2013.
- [21] Tom MATHES, Justus H. PIATER: Robust non-rigid object tracking using point distribution models. In In : Proc. of the British Machine Vision Conference, pages 849–858, 2005.
- [22] Tao Yang; Quan Pan; Jing Li; Li, S.Z., "Real-time multiple objects tracking with occlusion handling in dynamic scenes," *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol.1, no., pp.970,975 vol. 1, 20-25 June 2005.
- [23] Rabaud, V.; Belongie, S., "Counting Crowded Moving Objects," *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol.1, no., pp.705,711, 17-22 June 2006.
- [24] G. Farneback: "Two-Frame Motion Estimation Based on Polynomial Expansion", SCIA13, Lecture Notes in Computer Science, Vol. 2749, 363-370, 2003.
- [25] Vidéo course et marche Unusual crowd activity dataset of University of Minnesota, available from <http://mha.cs.umn.edu/movies/crowdactivity-all.avi>.
- [26] MEHRAN, Ramin, OYAMA, Akira, et SHAH, Mubarak. Abnormal crowd behavior detection using social force model. In : *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009. p. 935-942.